Is Cognitive Neuropsychology Plausible? The Perils of Sitting on a One-Legged Stool

Is Cognitive Neuropsychology Plausible? The Perils of Sitting on a One-Legged Stool

Stephen M. Kosslyn and James M. Intriligator
Harvard University

Abstract

■ We distinguish between *strong* and *weak* cognitive neuropsychology, with the former attempting to provide direct insights into the nature of information processing and the latter having the more modest goal of providing constraints on such theories. We argue that strong cognitive neuropsychology, although possible, is unlikely to succeed and that researchers will fare better by combining behavioral, computational, and neural investigations. Arguments offered by Caramazza (1992) in defense of strong neuropsychology are analyzed, and examples are offered to illustrate the power of alternative points of view. ■

INTRODUCTION

Is cognitive neuropsychology possible? Of course it is *possible;* nobody we know ever claimed otherwise. But that is not saying much—almost anything is possible. Rather than asking whether cognitive neuropsychology is possible, we should ask whether the goals of cognitive neuropsychology are plausible given the methods it uses. Cognitive neuropsychologists aim to understand "the structure of normal perceptual, motor, and cognitive processes" (pp. 80–81).[1] A theory of the structure of such information processing systems posits component processes (such as, in the case of reading, a letter-to-sound conversion process) and structures (such as a buffer that holds graphemic information temporarily), which are understood in part by specifying the properties of representations that reside in the structures and are produced and manipulated by processes. Cognitive neuropsychologists focus on observing selective deficits in behavior that occur after brain damage, and use the patterns of associated and dissociated deficits to draw inferences about the nature of normal human information processing.

Cognitive neuropsychologists focus almost entirely on patterns of functional deficits—unlike cognitive neuroscientists, they do not rely on facts about the brain when drawing their inferences about normal processing. We argue that patterns of deficits are simply too underconstraining to allow one to draw strong inferences about the underlying processing system. To be clear about our claims, we must distinguish between two variants of cognitive neuropsychology. *Weak* cognitive neuropsychology is the study of the behavior of normal and brain-damaged individuals to *constrain* theories of normal cognitive processing. On this view, the "principal or only aim [of cognitive neuropsychology] is to constrain theories of normal cognitive functioning through the analysis of acquired disorders of cognition" (p. 81). We brook no argument with this enterprise.[2] Unquestionably, to fully understand a working system one must understand the ways in which it can fail. Cognitive neuropsychological data clearly can serve to constrain theories of cognition and can be a source of inspiration for theorizing about the structure of normal cognition.[3]

In contrast, *strong* cognitive neuropsychology is the study of the behavior of normal and brain-damaged individuals with the goal of *inferring* the structure of normal cognitive processing. In this case, the goal is "to draw inferences about the structure of normal cognitive processes" (p. 80) and "to develop and evaluate theories of normal cognition" (p. 86).[4] Our argument is with the goals of strong cognitive neuropsychology. Although Caramazza (1992) often focuses on weak cognitive neuropsychology (which is easily defended) in his discussion, his use of data from patient NG—and much of the literature in the relevant journals—is an example of strong cognitive neuropsychology; the goal is to infer features of the processing system by observing behavioral disruptions following brain damage.[5]

The issue comes down to the following question: Can patterns of performance following brain damage in and of themselves reveal the nature of human (or other biological) information processing? Our argument is that strong cognitive neuropsychology is a discipline perched

perilously on a one-legged stool. It is certainly *possible* to sit on a one-legged stool, but its instability makes it too easy to shift positions. Why take the risk of falling off such a stool when it is so easy to add more legs, in the form of information about the neural substrate and explicit computational models of neural information processing? Which would you rather sit on, a one-legged or a three-legged stool?

Kosslyn and Van Kleeck (1990) argued that strong cognitive neuropsychology is unlikely to succeed, but noted that both neuropsychological data and normal behavioral data play a valuable role as constraints on theories.[6] Kosslyn and Van Kleeck's argument hinged in part on the view that the brain is a highly nonlinear, dynamic system—not a collection of isolated, discrete components. Neural subsystems are intimately interconnected, and hence many factors affect what a patient can and cannot accomplish following brain damage. For example, brain damage not only disrupts the processes carried out by damaged tissue, but also may disrupt connections, provide spurious inputs to (or "shock") remote intact tissue, result in a decrease in "activation" (and so more difficult tasks cannot be performed), and so on. These indirect disruptions may lead to various types of compensations and possibly the development of new processes, which change the behavior—sometimes producing the appearance of a deficit and at other times masking actual deficits.[7]

Much of Caramazza's (1992) discussion is a defense of weak cognitive neuropsychology, and we agree with this defense (as did Kosslyn & Van Kleeck, 1990). However, he also defends strong neuropsychology, repeatedly appealing to several general lines of argument in his discussion.[8] We avoid redundancy by not considering each of his points individually (many are closely related), but instead speak to his more general themes. We first consider these lines of defense, illustrating our points with several types of examples, and then apply our observations to the findings Caramazza (1992) uses to illustrate how patterns of behavior following brain damage can lead one to infer facts about cognitive function.[9]

LOGIC OF INFERENCE

The likelihood that strong cognitive neuropsychology can succeed depends on certain assumptions, many of which we find implausible. These assumptions are discussed in this section. In each case, we review Caramazza's (1992) position before offering our response.

The Fractionation and Transparency Assumptions

Caramazza writes that "Intuitively, we can assume that impaired performance (P^*) has the same relation to a model of the damaged cognitive system (M^*) as that of normal performance (P) to the normal cognitive system (M)" (p. 81). Caramazza writes that "there are various background assumptions that are supposed to motivate the use of particular performance measures . . . for inferring the functioning of the system(s) assumed to support cognitive performance, whether impaired ($P^* \rightarrow M^*$) or normal ($P \rightarrow M$)." (p. 81). Here it seems that one is to translate the arrow as indicating "is used to infer"—thus, ($P^* \rightarrow M^*$) can be translated as "impaired performance is used to infer a model of the damaged cognitive system." Finally, he defines L_i as a "functional" lesion.[10] Therefore, the translation table for Caramazza's formalism thus far is as follows:

impaired performance (P^*)
model of the damaged cognitive system (M^*)
normal performance (P)
normal cognitive system (M)
a functional lesion (L_i)
is used to infer ($\rightarrow$)

And his formalized argument runs as follows:

$$\frac{(P^* \rightarrow M^*) \qquad M^* = M + L_i}{P^* \rightarrow M + L_i}$$

This formalism helps to explicate two fundamental assumptions, both of which are necessary to infer components of the normal system from patterns of behavioral dysfunction. The *fractionation assumption* states that "brain damage can result in the selective impairment of components of cognitive processing" (Caramazza, 1984, p. 10). These components are defined by functional analyses. Thus, the term L_i refers to one processing component, L_j to another, and so on. The *transparency assumption* "essentially states that the cognitive system of a brain-damaged patient is fundamentally the same as that of a normal subject except for a 'local' modification of the system" (Caramazza, 1986, p. 52). Note that these assumptions are necessary only if one is engaged in *strong* cognitive neuropsychology, seeking to use the data to induce the underlying structure of the normal system. Neither assumption is necessary to use patterns of behavior in brain-damaged patients as *constraints* on theory (as will be illustrated shortly with the case of patient RV).

Response

The fractionation assumption has two parts. The first is the claim that brain damage can selectively impair different sorts of processing. The mere fact that the brain is not a homogeneous structure, with different regions having different input/output connections, suggests that this is true (see Chapter 2 of Kosslyn & Koenig, 1992). However, given the interconnectivity of neural structures (see below), damage rarely (if ever) will affect only a

single structure; the selectivity is a matter of degree. The second is the claim that damage affects components defined by "functional analyses." This is difficult to dispute, given that "functional analyses" can characterize many types of information processing. However, if functional analyses are based on common sense, linguistic theory, or some other behaviorally based inferences, there is no reason to accept this assumption. Components of behavior need not correspond to components of processing.

In addition, the transparency assumption seems patently false, implying that disrupted behavior reflects only the missing contribution of a damaged module. Brain function typically is nonlinear, and so models that assume simple additive effects are unlikely to be accurate (cf. McClelland, 1979). This nonlinearity arises from basic properties of the brain's anatomy and physiology. For example, the vast majority of connections between cortical areas are reciprocal, with connections running in each direction (Felleman & Van Essen, 1991; Van Essen, 1985). This architecture indicates that processing in multiple areas is intimately intertwined, and so damaging one area disrupts inputs to other areas. Damage affects the system as a whole, not isolated components. (These patterns of connections make sense if input is often noisy or ambiguous, and "cooperative computation" is used to overcome these problems; for a discussion, see Kosslyn & Koenig, 1992.)

Furthermore, behavior following brain damage may reflect not the effects of a missing or impaired module, but rather the fact that processes are not interacting in the usual way. To be concrete, consider the findings of Kosslyn, McPeek, Daly, Alpert, and Caviness (1991c) on patient RV, who had a lesion in the left frontal lobe. In an MRI scan, it appeared likely that this lesion disrupted the inferior longitudinal fasciculus, which would have de-enervated posterior regions of the left hemisphere. And in fact, a PET scan revealed a large region of hypometabolism in the left occipitotemporal area. This anatomical and physiological information suggested that RV might have a deficit in encoding "object properties," such as shape (which are known to be encoded in the "ventral," temporal lobe-based system; see Maunsell & Newsome, 1987; Mishkin, Ungerleider, & Macko, 1983, for reviews). However, this damage did not affect, directly or indirectly, the areas known to be involved in encoding "spatial properties," such as location (which are thought to be encoded in the "dorsal," parietal-lobe based system). In fact, RV showed a visual deficit: Unlike control subjects, he required progressively more time to encode more complex shapes that were formed by filling in cells of 4×5 grids. Not only did he require more time to determine whether two sequentially presented stimuli were the same or different when they incorporated more cells, but also required more time simply to decide whether an X mark fell on or off the more complex shapes. The deficit in both tasks disappeared when the internal grid lines were removed.

At first glance, it might be tempting to suspect that RV had a deficit in encoding visual "features," such as lines and vertices, and the grid lines simply overloaded this impaired module. Such an inference would follow from the fractionation and transparency assumptions. But consider how such an inference fares in light of additional findings: When the grid lines were removed, RV did not require less time overall to match sequentially presented patterns than he did when they were in grids. Furthermore, when random noise elements were placed over patterns that were not in grids, he did *not* require more time for more complex patterns; the lines had to form an orderly array of grid cells to define sets of locations before the complexity effect was observed. Neither result makes sense if RV simply had an impaired "feature encoding" module.

In contrast, these results are as expected if the grids defined sets of locations (the cells), and the intact parietal lobe-based spatial encoding mechanisms encoded shapes as sets of locations of filled cells—and more locations require more time to encode. There was no increase in time with increasing stimulus complexity for a control group, which suggests that in normal people this location coding process requires more time than simply encoding a shape, and so its outputs do not end up being used to perform shape comparison tasks. However, in RV, the damage slowed down his temporal lobe-based shape encoding mechanisms, which caused the output of the slower but still effective spatial encoding system to be used when a grid was available.

Other accounts of these findings are possible, but this one has the advantage of being consistent with facts about the brain and computational analyses (see Chapter 3 of Kosslyn & Koenig, 1992). Our point here is not to argue for this particular explanation, but rather to illustrate how a different perspective—viewing a patient's deficit as reflecting an alteration of the "ecological balance" of the processing system—can lead one to collect data that are not easily explained as the missing contribution of one or more individual components. We are not trying to engage in strong cognitive neuropsychology, but rather are treating these results as constraints on theorizing: Whatever the ultimate account of these findings, they suggest that deficits should be understood as alterations of a system of interacting components, not as isolated, local modifications of the normal system, which otherwise continues to operate normally. The intact components do not necessarily contribute to behavior normally, as implied by Caramazza's formalism.

In addition, there is evidence that the brain at least sometimes actually reorganizes following brain damage (e.g., Jacobs & Donoghue, 1991; Merzenich, Kaas, Wall, Nelson, Sur, & Felleman, 1983; Merzenich, Nelson, Stryker, Cynader, Schoppman, & Zook, 1984; Pons, Garraghty, Ommaya, Kaas, Taub, & Mishkin, 1991). Caramazza decries this possibility, suggesting that if it is true then it will be impossible to engage in (strong) cognitive neuro-

ory, which directs them to focus on specific aspects of the phenomena. They know that these measurements will allow them to answer certain questions because they can rely on a host of background assumptions that are embedded in a rigorous and well-articulated theoretical framework. In cognitive neuropsychology the goal is to construct the very sort of theory that is a prerequisite for using the cloud chamber. In our case, we cannot be certain whether overall speed, speed of initiating the response, force of responding, overall accuracy, variability in accuracy over trials, and so on are the appropriate measures (e.g., see Abrams & Balota, 1991). If some other measure were used, the "intact" ability might not appear so intact. For example, if one only examines errors, one might give a clean bill of health to a patient who takes 10 times longer than normal to respond. To be concrete, if Caramazza and Hillis (1991) had measured response times, they might have found that their patients required abnormal times for nouns—which would have challenged the inference that verbs are represented in a distinct structure.

Data Underdetermine Theory

Cognitive neuropsychologists acknowledge that there are many possible theories or hypotheses that can account for observed behavioral data. However, one could argue that this is not a special feature of cognitive neuropsychology. Rather, there are always alternate theories that could account for *any* empirical observations, and the induction from the behavior of brain-damaged individuals to a correct theory of normal cognitive processing is no more problematic than the induction of any theory from any type of observation.

Response

Although the induction of theory from data is always problematic and underdetermined, this is a matter of degree. It is one thing to be working within a well-established framework in which questions can be cast and observations interpreted, and quite something else to be working in a field where the basic structure of theories is at issue. (Consider the amount that could have been learned from cloud chambers—let alone response times to visual stimuli—in the third century A.D.) In some branches of physics, it is difficult to produce even one plausible competing account for an empirical finding.

In contrast, it is easy to generate many alternative accounts for results in cognitive neuropsychology, and these accounts may rest on assumptions at different levels of analysis (ranging from aspects of the task or instructions to the amount of effort to produce the response). One advantage of trying to understand the nature of information processing and brain mechanisms at the same time is that brain mechanisms are constrained by the laws of physics, and so alternative accounts of mechanisms are difficult to formulate. For example, it is difficult to produce more than one credible explanation of how a neuron fires.

Caramazza's argument is a little like pointing out that one can slip and break one's neck in the shower or by walking on a tightrope, and thereby concluding that both practices are equally dangerous. Although both risks exist, they differ by a matter of degree. In this analogy, physics is like taking a shower: there is only a small risk that one will come to a bad end because the danger is highly restricted and easy to control.

Descriptions of Deficits Are Theory-Related

The way we characterize a deficit depends on our theory, but this is true for any scientific observation and "we are still led to ask whether the implications of this fact are particularly problematic for cognitive neuropsychology." (p. 87)

Response

Our response parallels our previous one; it is a matter of degree. To the extent that a rigorous theoretical framework does not already exist, one has many degrees of freedom when describing data. Thus, although this factor affects all sciences, it is particularly troublesome when the essential elements of a theory are in dispute.

THE PRACTICAL VERSUS THE POSSIBLE

Much of medicine is based on pragmatic considerations. If a drug works, it is used—even if its mechanism is not understood, or if theory suggests that it should not work. Caramazza argues that (strong) cognitive neuropsychology seems to be working, so possible logical objections should be put aside. We consider these claims in this section.

Pragmatics as a Guide

Caramazza recommends that we should not be guided by logic alone, but rather: "The justification is strictly pragmatic: we are justified in using the performance of brain-damaged subjects to infer the structure of normal cognition if, *in practice,* these inferences lead to significant insight into the nature of normal cognitive processing" (p. 82). He further claims that it is "an empirical matter that cannot be decided by logic alone. Consequently, the justification for undertaking the enterprise must ultimately be based on pragmatic considerations: that is, on considerations about the productivity of the enterprise in generating significant insights into the problems it has chosen to address" (p. 89). In other words, even though there is no ironclad logical reason that performance measures of brain-damaged individuals

should directly implicate theories of normal cognition, if these performance measures seem to work in practice then we should accept them.

Response

How does one decide whether this strong cognitive neuropsychology enterprise is "working"? It is easy to argue that very few insights about normal cognition have emerged solely from studies of brain-damaged patients. Indeed, without explicit computational models, or clear relations to neuroanatomy or neurophysiology, it is difficult to know how to determine when progress has been made; it is too easy to wave one's arms around and rely on the vagueness of natural language when explaining a finding.

Coherent Patterns of Performance

Even though it is not logically necessary, coherent patterns of performance following brain damage appear to offer insight into the structure of cognitive processing. For example, when deficits tend to cluster, this seems to suggest that they share a common underlying processing component. "The guarantees we have are strictly pragmatic in nature. They spring from the fact that the performance of brain-damaged subjects appears to be patterned in a coherent fashion, and investigation of these patterns of performance *seems* to lead to interesting insights about normal cognitions" (p. 85). Caramazza notes that in clinical neurology it "was repeatedly observed that, with notable frequency, brain damage resulted in highly specific cognitive, perceptual, and motor deficits" (p. 89).

Response

We do not dispute that behavior can be disrupted in orderly ways. The issue is whether one can infer the underlying functional bases for such patterns of disruption solely by observing behavior. Even if tasks that are impaired together share a common processing component, this component might bear an abstract relation to the observed behavioral deficit. Neural network computer simulations have shown that tasks can be accomplished using representations that are not intuitively obvious. For example, Lehky and Sejnowski (1988a,b) trained a network to extract shape from variations in the shading of a surface, and found that it developed "end-stopped" hidden units; presumably, if these units were damaged, the network would be impaired at extracting shape from shading. Armed with this hypothesis, one could look for a deficit in detecting termini of lines in patients who have trouble deriving shape from shading. But one would probably never infer such an "implausible" mechanism on the basis of the behavioral data alone.

Moreover, the fact that deficits sometimes cluster together could occur if the tasks require similar amounts of "activation," are indirectly affected by spurious inputs from another region, or require additional blood flow either to or through a single damaged locus. Or it could reflect the complexity of the instructions, the effort required to generate a response, and so forth (see Kosslyn & Koenig, 1992).

THE AUTONOMY OF FUNCTIONAL ANALYSES

Strong cognitive neuropsychology rests on the assumption that one can infer the nature of cognitive functions independently of considerations about the brain or of detailed models of computational systems. We are skeptical, for the reasons noted below.

Neural Reality

Cognitive neuropsychologists do not focus on how functions are realized in neural hardware. Furthermore, Caramazza argues that we should not require theories of cognitive processing to be neurally accurate because if we were to adopt this "stringent criterion for determining the level of interest in a cognitive theory, we would have to consider as of 'little interest' the vast majority of cognitive neuropsychological research, seeing as most of it is concerned with complex cognitive functions (e.g., language) for which at this time there is not much detailed information at the neural level" (p. 87).

Response

A given behavior is produced by one sequence of information processing, and not others. There is a "fact of the matter;" some theories are correct, and others are incorrect. The demonstration that a theory is "computationally adequate" is a necessary but not sufficient measure of its veracity; we also want evidence that those processes are actually carried out by the brain. If the theories of cognitive processing that are formulated by cognitive neuropsychologists do not reflect the way the brain works, then they are of little value for cognitive science or neuroscience (although they may be of interest in artificial intelligence). In many cases, we do not yet know whether the brain embodies the distinctions of specific theories, but in our view this is not simply icing on the cake: Researchers should seek to determine the neurological reality of their putative functional distinctions. The function being described is, after all, the function of the brain, not of the big toe or some other organ.

Thus, drawing inferences about function without regard to the brain will, at best, provide only some of the information needed to evaluate a theory of human (or other biological) information processing. Our point is simple: The hypotheses and theories that are generated by cognitive neuropsychologists are of little interest if

they are incorrect, and one cannot evaluate the theories rigorously solely by considering behavior.

Computational Models

It is not clear to some "how reliance on explicit theory would overcome the putative defects of neuropsychological research" (p. 91). Indeed, "if it were to turn out that brain damage does in fact create 'new functions,' then, no matter how detailed our cognitive theories might be, the performance of brain-damaged subjects could not be of use in constraining normal theory" (pp. 91–92).

Response

The brain is a dynamic system, and real-time interactions among component processes can be modeled on a computer. Analyses of how to build a model that can mimic specific behavior is one source of hypotheses about processing, and actually building computational models can help one to discover the empirical implications of one's ideas—which are not always clear when one is dealing with complex nonlinear systems; static, linear formalisms are likely to have limited use in understanding brain function. Computational models are particularly useful because they allow one to simulate complex properties of the brain, and use these properties as constraints on theories of information processing.[11] For example, Kosslyn, Flynn, Amsterdam, and Wang (1990) implemented a model of visual object identification that is organized in terms of the major pathways of high-level vision. This model can be damaged, allowing one to anticipate effects of disconnections, compensations, and so forth. As noted earlier, predictions from computational models can be very nonintuitive.

In addition, computational models can help one to discern what sorts of new functions could emerge following damage; such functions do not magically appear out of whole cloth, but arise within the context of the surviving aspects of the system (cf. Pearson et al., 1987). Computational models can help one to understand what sorts of new functions might arise following specific types of damage—and so can generate empirically testable hypotheses. Indeed, a weak cognitive neuropsychological approach is particularly useful when one has a computer model: If the model cannot account for relevant observed phenomena, it must be ruled out.[12]

THE LIMITS OF STRONG COGNITIVE NEUROPSYCHOLOGY: AN EXAMPLE

In discussing the case of patient NG, Caramazza avers that although Kosslyn and Van Kleeck's criticisms may sound plausible in the abstract, they carry no force when confronted with actual data. Caramazza's discussion of NG is a good example of the strong cognitive neuropsychological stance; he clearly wants to draw inferences about an underlying processing system in normal people based on the patient's performance per se. Our objections to this practice are illustrated by his use of these data.

Patient NG was a left-handed woman who apparently had a lesion of the left parietal white matter and the left anterior basal ganglia, adjacent to the head of the caudate. She neglected (ignored) the ends of words no matter how they were oriented in space: If the words were vertical, she ignored their bottoms; if they were mirror-reversed, she ignored letters at the left side, and so on. She also ignored the final letters of words when they were spelled aloud. Similarly, she tended to make errors at the end of words when spelling them aloud or writing them. From these and similar results, Caramazza draws four conclusions, none of which necessarily follows from the data he presents.

We begin with Caramazza's second conclusion, which lies at the heart of his claims, namely that the findings rule out deficits at "retinocentric and stimulus-centered levels of representation" and instead demonstrate the existence of a "word-centered" representation. But this conclusion does not necessarily follow, for it is possible that the problem has to do not with the *representation* of words or other stimuli, but rather with the *processing* of this information. Caramazza wonders how it would matter whether a function is implemented in a small group of nearby neurons or distributed widely. One answer is that if we assume that processes are implemented by widely distributed neurons, degraded performance—not the all-or-none presence of a component—should be the rule following brain damage because (as Kosslyn and Van Kleeck noted) a lesion is unlikely to obliterate all of the relevant neurons. With this in mind, first consider the fact that the left parietal lobe was damaged. The parietal lobes are known to be critically involved in computing spatial properties of stimuli and have a critical role in directing attention (for reviews see Andersen, 1987; Posner & Petersen, 1990; Ungerleider & Mishkin, 1982). Humans apparently encode each letter of a word separately when reading, scanning from the beginning of the word to the end (Just & Carpenter, 1980, 1987). The damage may have impaired NG's ability to estimate distances properly—leading her to underestimate the amount of scanning that is required to encode an entire word. If so, then—like normal subjects—she encodes the letters one at a time from the beginning of the word, but fails to scan far enough to encode them all when reading the whole word. Such scanning of the overall pattern is not necessary to read the letters one at a time (which she could do). The data suggest that she underestimates a relatively constant percentage of the length, not a fixed amount. This scanning operation would occur over a viewer-

centered representation, such as those that exist within the retinotopically mapped areas of the occipital lobe (see Felleman & Van Essen, 1991).

But what about NG's failure to decode orally spelled words? If NG performs this task by visualizing the word as it is spelled, then these results are easily explained. Kosslyn and Koenig (1992) review much data indicating that visual mental imagery shares processing mechanisms with visual perception. Thus, the scanning deficit evident in perception would also disrupt her ability to scan visualized words. Kosslyn, Alpert, Maljkovic, Weiss, Thompson, Hamilton, and Chabris (1991a) used PET scanning to study the brain bases of visual mental imagery, and found that primary visual cortex is selectively activated during imagery. The fact that this area is retinotopically mapped in humans (Fox et al., 1986) is consistent with our view that the representation underlying NG's performance was not word-centered.

Finally, what about the fact that NG's written and oral spelling also revealed neglect of the right halves of the words? One account of this finding is that NG visualized the words prior to writing or orally spelling them. Normal people report doing this for "difficult" words, presumably in an effort to reconstruct information that is not strongly represented in memory. NG has brain damage, and so she may visualize words in general—even ones normal people would not call "difficult"—prior to spelling them.

Alternatively, another account hinges on the observation that NG's striatum apparently was compromised. If we can generalize from the macaque monkey to humans, this structure plays a critical role in habitual behavior (e.g., see Mishkin & Appenzeller, 1987). One aspect of reading may be a habit, namely estimating how many letters must be scanned across before beginning to read a word; other sequential tasks are "set up" in advance, before the process is actually initiated (e.g., see Sternberg, Monsell, Knoll, & Wright, 1978). Thus, the scanning problem could arise from a disruption of this "habit" system.

As yet another alternative, NG's problems may reflect a decrease in the "capacity" of a process that allocates "effort" for performing sequential tasks. It is possible that there is a process that uses preattentive information to allocate capacity for shifting attention. If this process were realized in a small number of neurons that were widely distributed, it might become degraded but not entirely dysfunctional. Hence, it would simply fail to allocate enough processing capacity (however defined), and scanning would fall short. If such a process were shown to be highly localized, or to involve many, redundant neurons, this conjecture would seem implausible.

Caramazza and Hillis' (1990a,b) findings are difficult to evaluate for a number of reasons. First and foremost, they apparently did not collect response times. It is very difficult to interpret error rates without also knowing how long the subject needed to respond, if only because it is possible that there were speed/accuracy tradeoffs: NG may have responded more quickly than age-matched control subjects with lesions of similar size in other areas (which would control for the general slowing typically observed following brain damage), perhaps because she was anxious about being tested, was trying to please the experimenter, failed to estimate properly the amount of necessary processing before responding, and so on. If so, then she may not have allowed herself enough time to scan across the entire word before responding. Second, from the perspective of our accounts, it would have been useful to measure the time NG needed to speak or write each letter of a word. If progressively more time were taken toward the end of the word, this might suggest that NG "ran out of steam" too soon, not properly estimating the difficulty of the task. It would also be useful to collect such data from control subjects; it is possible that NG simply has an exaggerated case of a condition that appears commonly as the brain degrades with age. Moreover, if she required more time to read words when the letters were spread apart, this would be consistent with the scanning notion developed above. It is clear that Caramazza's theoretical preconceptions led him to collect some data and not others, and the available data are consistent with numerous alternative accounts.

Knowledge of how function is implemented in the brain could play a critical role in discriminating among the various alternative hypotheses we have offered: Once we know something about the function carried out by a particular part of the brain, we can apply that knowledge to understanding the deficits of patients with lesions in that area. In addition, depending on how function is implemented, an account that posits that a lesion has directly affected a single function is more or less plausible. Information about how function is implemented in neural tissue is invaluable if we are to distinguish among the many possible alternative accounts for any set of behavioral dysfunctions following brain damage.

Now let us consider the other three conclusions Caramazza draws on the basis of behavioral dysfunctions following brain damage.

Conclusion 1 is that NG's impairment is at "a level of processing that specifies the identity and order of graphemes (abstract letter identities) and not specific letter shapes" (pp. 83–84). The alternatives above all posit that viewer-centered representations of specific shapes are used. Hence, this conclusion does not necessarily follow.

Conclusion 3 hinges on the fact that other patients have been studied who always neglect stimuli in the left visual field, and so neglect different parts of words when they are presented normally than when they are presented mirror-reversed (e.g., see Behrmann, Moscovitch, Black, & Mozer, 1990). Caramazza wants to conclude that there is therefore "a distinction between a canonical, word-centered" and a "stimulus-centered" representation

of words. However, these latter subjects neglect the left visual field, and the deficit is not specific to words. But more importantly, these findings may simply suggest that scanning mechanisms can be disrupted in more than one way (see Chapter 5 of Kosslyn & Koenig, 1992).

Finally, *Conclusion 4* is that NG's deficit concerns the "right *half* of a grapheme representation" (p. 84). We note in Figure 1 of Caramazza's article that this patient also neglected the right half of some types of objects, which suggests that the deficit was not restricted to the right side of graphemic representations per se.

Caramazza then goes on to "summarize" his conclusions in a model of the component processes used in normal word recognition. This model posits three independent levels of representation. The first level, a *feature map,* consists of a "retinocentric description of the edges in a retinally projected image." It is unclear how the data led to this conclusion, and we suspect that he is borrowing ideas from other types of research (computational modeling and neurophysiology) to formulate this idea. The second level, the *letter-shape map,* appears to be a viewer-centered description of the shapes and spatial relations among letters. Although he claims that this representation is analogous to Marr's 2.5-D sketch, it is unclear whether this claim should be taken at face value: Marr's representation did not have explicit representations of edge boundaries; rather, it was a depth map, which used a "pin cushion" representation to make explicit properties of surfaces. Moreover, Caramazza appears to posit explicit representations of the spatial relations among letters, which also was not a feature of Marr's 2.5-D representations. Indeed, there is good evidence that this level of representation does not specify spatial relations explicitly, given that monkeys who have intact occipital and temporal lobes (the probable loci of this functional representation) but missing parietal lobes have impaired representation of spatial relations (e.g., see Ungerleider & Mishkin, 1982). Furthermore, we have seen no evidence for a distinct representation for letter shapes per se; indeed, Caramazza's own logic of inference ($P^* \rightarrow M + L_i$) seems to suggest that a single nonlinguistic processing component has been damaged, given that the patient's deficits also affect nonlinguistic stimuli. (He could suggest that there are two functional lesions here, but we see no grounds for this inference.) Finally, consider the third level, the *grapheme description.* Caramazza says, "In order to account for the results obtained with NG we must assume the hypothesized distinction between the latter two levels of representations: NG has a spatially specific deficit at the level of the grapheme description and not at the level of the letter-shape map" (p. 85). As we have seen, we must assume no such thing.

We found it rather striking that immediately after this exercise in strong cognitive neuropsychology, Caramazza asks whether there are reasons for "excluding the performance of these subjects from the range of facts that may be relevant for the purpose of deciding among competing accounts of the process of word recognition" (p. 85). Of course there are not; this is proper weak cognitive neuropsychology. But it is an error to confuse the two enterprises, as Caramazza appears to do throughout his article. We are skeptical about the claim that behavioral data from brain-damaged patients alone can implicate components of a processing system, not that they are important constraints on all theories.[13]

CONCLUSIONS

Caramazza argues that "behavioral observations of brain-damaged subjects can stand on their own in the development of a meaningful cognitive science . . . developments in cognitive science concerning the computational structure of cognitive processes can proceed independently of neuroanatomical observations" (p. 85). We agree. Of course behavioral observations of brain-damaged subjects can be conducted without regard to neuroanatomy, and such observations will provide useful constraints on theory. But these observations, standing on their own, are not likely to implicate a correct theory of information processing. Strong cognitive neuropsychology is a nineteenth-century endeavor, and the reasons it failed then are still with us today. To the extent that cognitive neuropsychology is succeeding, it is because theorists are using computational ideas or are discovering surprising phenomena. When phenomena defy common sense, they often imply that conventional ways of conceptualizing a problem or theoretical assumptions are incorrect, which is always useful. Such findings place constraints on all theories, which must now account for these nonintuitive results. But, as useful as they are, such findings do not directly reveal the nature of the underlying mechanisms.

Our argument is simple: Why try to sit on a one-legged stool when one can use a three-legged one? In addition to behavioral data, computational modeling and neural constraints can play a critical role in helping one to formulate and test theories. It is difficult for us to see how one could disagree with this observation. And in fact, in spite of all of his arguments to the contrary, Caramazza himself writes, "it is amply evident that such information [anatomical and neurophysiological] is fundamental for any effort directed at developing and constraining theories of the functional organization of the brain" (p. 92). Moreover, "a nontrivial theory of the 'functional organization of the brain' will be a theory of the neural implementation of specific cognitive processes" (p. 93). We couldn't have said it better ourselves.

Acknowledgments

Preparation of this article was supported by AFOSR Grant 91-0100 and NSF Grant BNS 90-09619. The second author was supported by an NSF Graduate Fellowship. We thank Christo-

pher Chabris for his careful reading of an earlier draft and his editorial assistance, and Lisa Shin for valuable comments and help in classifying articles in the journal *Cognitive Neuropsychology.*

Notes

1. Unless otherwise noted, all quotations are from Caramazza (1992).
2. Kosslyn and Van Kleeck stated that "studying the effects of brain damage is without question one source of evidence for a theory of information processing" (p. 400).
3. Caramazza accuses Kosslyn and Van Kleeck of concluding "that the study of brain-damaged subjects for the purpose of constraining theories of normal cognitive processing is doomed to failure" (p. 85). However, Kosslyn and Van Kleeck never make any such claim. The nearest claim they make is that "it is virtually impossible to induce a correct theory of information processing simply by observing patterns of deficits following brain damage" (p. 391). Constraining theories of normal cognition is a useful goal for cognitive neuropsychologists.
4. Caramazza refers to the "functional organization of the brain" (p. 81), but it is not clear what he means by this phrase. Given the focus of cognitive neuropsychology on functional organization, it might be both less confusing and more accurate to say the "functional organization of behavior"—which is what is actually being studied. It is useful to be clear on the distinction between brain and behavior and the distinction between implementation and function.
5. Caramazza writes, "Whatever may be [Kosslyn and Van Kleeck's] motivation for ascribing a naive inductivist view of science to cognitive neuropsychologists, it should be apparent that there is nothing intrinsic to cognitive neuropsychology that requires that one adopt this position" (p. 86). Their motivation was based on reading journals such as *Cognitive Neuropsychology, Brain and Cognition,* and *Brain and Language*—all of which include a substantial number of articles that adopt the strong cognitive neuropsychology approach. For example, we selected a volume of *Cognitive Neuropsychology* at random (1987), and counted the number of articles that adopted a strong cognitive neuropsychology approach. We evaluated each of the 10 articles that reported investigations of brain-damaged subjects, using a conservative criterion for "strong" cognitive neuropsychology, and found that 5 clearly fell in this category. We agree that there is nothing intrinsic to the research that requires this approach, but it certainly characterizes a major trend in the field.
6. Caramazza accuses Kosslyn and Van Kleeck of making a "sweeping condemnation of cognitive neuropsychological research" (p. 85), but this is a misreading. Kosslyn and Van Kleeck argue against strong cognitive neuropsychology, and point out that this approach is unlikely to allow one to infer a correct theory of human information processing. This was not meant to degrade the contribution that can be made by cognitive neuropsychologists.
7. Note that the only way in which deficit data can serve a role in theorizing is by being couched in terms of deficits, and hence a characterization of normal operation must *precede* deficit research. Without such a characterization of normal functioning it is not clear what one should consider a "deficit."
8. Caramazza claims that Kosslyn and Van Kleeck raised two kinds of points, *in principle* and *in practice.* This is a misreading: All of the points were intended to be *in principle;* more careful measurements will not cure problems of inference.
9. Caramazza claims that the study of abnormal behavior by Broca, Charcot, Jackson, and Wernicke "led to the first explicit and empirically defensible claims about the relationship between neuroanatomy and cognitive processes" (p. 80). These goals are more in line with cognitive neuroscience than with cognitive neuropsychology; neuroanatomy almost never enters into discussions of cognitive neuropsychology. Furthermore, relatively little that was discovered by these pioneers is accepted as entirely accurate today.
10. The concept of a functional lesion appears at times to be conflated with the concept of a structural (physical) lesion. For example, Caramazza refers to "brain-damage on the cognitive system" and "functional lesion to the cognitive system" (p. 82). We know what brain damage is, and Caramazza usually uses "cognitive system" to refer to a functional description of the system (as in the second quotation), but we must take care not to assume that the two can be discussed with interchangeable terms.
11. We are not suggesting that all research with patients requires the use of computer simulation models or that these models are always a good idea. Computational models must incorporate many arbitrary details if the domain is not reasonably well understood. In the initial phases of research, it often is better to investigate issues, trying to discriminate among alternative positions (see Kosslyn, 1980). If the issues concern the existence of specific processing components, one source of useful evidence is the existence of selective deficits. However, given the complexity of the issues and loose linkage from data to theory, no single source of evidence is compelling. Hence, even here we argue that a convergent, interdisciplinary approach is likely to be most useful.
12. Caramazza inaccurately ascribes to Kosslyn and Van Kleeck the position that "neuropsychological investigations that are not explicitly guided by neuroanatomical or neurophysiological considerations cannot lead to meaningful conclusions" (p. 85) and "no meaningful conclusions can be reached in this area" (p. 87). One problem is with the term "meaningful": there is no doubt that neuropsychological investigations can lead to meaningful conclusions about the *constraints* that a theory must respect. Another problem is with the word "lead": we reject strong cognitive neuropsychology, but advocate weak cognitive neuropsychology—which may play a role in leading to conclusions, but does not do so in isolation. In addition, Caramazza writes that "Kosslyn and Van Kleeck claim that the cognitive neuropsychologist's interest in neuropsychological data is principally motivated by his/her disaffection with the methods of cognitive psychology" (p. 86). However, the claim that was actually made is that "part of the appeal of neuropsychology derives from disillusionment with the strictly behavioral approach." (Kosslyn & Van Kleeck, 1990, p. 390).
13. It is of interest that Caramazza chose to describe a patient with visual/spatial problems, rather than one of the patients he has studied with language deficits. Physiological studies of vision and computational modeling give weight to his first conclusion. If Caramazza had considered these other sources of support on an equal footing with the patient's performance, this conclusion could have been defended much more strongly. Moreover, thinking about properties of computational systems and functional anatomy might have led him to collect additional data (including response times) when he and Hillis studied NG, and perhaps to arrive at different conclusions. If Caramazza had presented other cases, such as the patients with verb production deficits noted above, we would have had an even clearer example of strong cognitive neuropsychology at work.

Reprint requests should be sent to S. M. Kosslyn, Harvard University, 1236 William James Hall, 33 Kirkland Street, Cambridge, MA 02138.

REFERENCES

Abrams, R. A., & Balota, D. A. (1991). Mental chronometry: Beyond reaction time. *Psychological Science, 2,* 153–157.

Andersen, R. A. (1987). The role of the inferior parietal lobule in spatial perception and visual-motor integration. In F. Plum & V. B. Mountcastle (Eds.), *Handbook of physiology: The nervous system. Vol. 5.* Bethesda: American Physiological Society.

Behrmann, M., Moscovitch, M., Black, S. E., & Mozer, M. C. (1990). Perceptual and conceptual mechanisms in neglect dyslexia: Two contrasting case studies. *Brain, 113,* 1163–1183.

Caramazza, A. (1984). The logic of neuropsychological research and the problem of patient classification in aphasia. *Brain and Language, 21,* 9–20.

Caramazza, A. (1986). On drawing inferences about the structure of normal cognitive systems from the analysis of impaired performance: The case for single-patient studies. *Brain and Cognition, 5,* 41–66.

Caramazza, A. (1992). Is cognitive neuropsychology possible? *Journal of Cognitive Neuroscience, 4,* 80–95.

Caramazza, A., and Hillis, A. (1990a). Spatial representation of words in the brain implied by studies of a unilateral neglect patient. *Nature, 346,* 267–269.

Caramazza, A., & Hillis, A. (1990b). Levels of representation, coordinate frames, and unilateral neglect. *Cognitive Neuropsychology, 7,* 391–445.

Caramazza, A., & Hillis, A. E. (1991). Lexical organization of nouns and verbs in the brain. *Nature, 349,* 788–790.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in primate cerebral cortex. *Cerebral Cortex, 1,* 1–47.

Fox, P. T., Mintun, M. A., Raichle, M. E., Meizen, F. M., Allman, J. M., & Van Essen, D. C. (1986). Mapping human visual cortex with positron emission tomography. *Nature (London), 323,* 806–809.

Jacobs, K. M., and Donoghue, J. P. (1991). Reshaping the cortical motor map by unmasking latent intracortical connections. *Science, 251,* 944–947.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87,* 329–354.

Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension.* Newton, MA: Allyn and Bacon.

Kosslyn, S. M., Alpert, N. M., Maljkovic, V., Weiss, S., Thompson, W., Hamilton, S. E., and Chabris, C. F. (1991a). *Visual mental imagery activates primary visual cortex.* Harvard University manuscript.

Kosslyn, S. M., Chabris, C. F., Marsolek, C. J., & Koenig, O. (1991b). Categorical versus coordinate spatial representations: Computational analyses and computer simulations. *Journal of Experimental Psychology: Human Perception and Performance,* in press.

Kosslyn, S. M., Flynn, R. A., Amsterdam, J. B., & Wang, G. (1990). Components of high-level vision: A cognitive neuroscience analysis and accounts of neurological syndromes. *Cognition, 34,* 203–277.

Kosslyn, S. M., and Koenig, O. (1992). *Wet mind: The new cognitive neuroscience.* New York: Free Press.

Kosslyn, S. M., McPeek, R. M., Daly, P. F., Alpert, N. M., & Caviness, V. S. (1991c). Using locations to store shape: An indirect effect of a remote lesion. Harvard University manuscript.

Kosslyn, S. M., and Van Kleeck, M. H. (1990). Broken brains and normal minds: Why Humpty-Dumpty needs a skeleton. In E. L. Schwartz (Ed.), *Computational Neuroscience.* Cambridge, MA: MIT Press.

Lehky, S. R., & Sejnowski, T. J. (1988a). Network model of shape-from-shading: Neural function arises from both receptive and projective fields. *Nature (London), 333,* 452–454.

Lehky, S. R., & Sejnowski, T. J. (1988b). Neural network model for the cortical representation of surface curvature from images of shaded surfaces. In J. S. Lund (Ed.), *Sensory processing in the mammalian brain.* Oxford: Oxford University Press.

Maunsell, J. H. R., & Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. *Annual Review of Neuroscience, 10,* 363–401.

McClelland, J. L. (1979). On the time-relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86,* 287–330.

Merzenich, M. M., Kaas, J. H., Wall, J. T., Nelson, R. J., Sur, M., & Felleman, D. J. (1983). Topographic reorganization of somatosensory cortical areas 3b and 1 in adult monkeys following restricted deafferentation. *Neuroscience, 8,* 33–55.

Merzenich, M. M., Nelson, R. J., Stryker, M. P., Cynader, M., Schoppman, A., & Zook, J. M. (1984). Somatosensory cortical map changes following digit amputation in adult monkeys. *Journal of Comparative Neurology, 224,* 591–605.

Mishkin, M., & Appenzeller, T. (1987). The anatomy of memory. *Scientific American, 256,* 80–89.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in NeuroSciences, 6*(11), 414–417.

Pearson, J. C., Finkel, L. H., & Edelman, G. M. (1987). Plasticity in the organization of adult cerebral cortical maps: A computer simulation based on neuronal group selection. *Journal of Neuroscience, 7,* 4209–4223.

Pons, T. P., Garraghty, P. E., Ommaya, A. K., Kaas, J. H., Taub, E., & Mishkin, M. (1991). Massive cortical reorganization after sensory deafferentation in adult macaques. *Science, 252,* 1857–1860.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. In W. M. Cowan, E. M. Shooter, C. F. Stevens, & R. F. Thompson (Eds.), *Annual Review of Neuroscience* (pp. 25–42). Palo Alto, CA: Annual Reviews, Inc.

Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). The latency and duration of rapid movement sequences: comparisons of speech and typewriting. In G. E. Stelmach (Ed.), *Information Processing and Learning* (pp. 117–152). New York: Academic Press.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of Visual Behavior.* Cambridge: MIT Press.

Van Essen, D. (1985). Functional organization of primate visual cortex. In A. Peters & E. G. Jones (Eds.), *Cerebral Cortex.* New York: Plenum Press.

This article has been cited by:

1. Matthew D. Lieberman, Grace Y. Chang, Joan Chiao, Susan Y. Bookheimer, Barbara J. Knowlton. 2004. An Event-related fMRI Study of Artificial Grammar Learning in a Balanced Chunk Strength DesignAn Event-related fMRI Study of Artificial Grammar Learning in a Balanced Chunk Strength Design. *Journal of Cognitive Neuroscience* **16**:3, 427-438. [Abstract] [PDF] [PDF Plus]
2. Kevin N. Ochsner, Matthew D. Lieberman. 2001. The emergence of social cognitive neuroscience. *American Psychologist* **56**:9, 717-734. [CrossRef]
3. Barry Horwitz. 1994. Data analysis paradigms for metabolic-flow data: Combining neural modeling and functional neuroimaging. *Human Brain Mapping* **2**:1-2, 112-122. [CrossRef]
4. Wim E. Crusio, Herbert Schwegler, Ingrid Brust. 1993. Covariations Between Hippocampal Mossy Fibres and Working and Reference Memory in Spatial and Non-spatial Radial Maze Tasks in Mice. *European Journal of Neuroscience* **5**:10, 1413-1420. [CrossRef]